Characterization of high-rate large-sized flows

Tian Jin*, Chris Tracy†, Malathi Veeraraghavan*
* University of Virginia
Charlottesville, VA 22904–4743
Email: {tj3sr,mvee}@virginia.edu
† Energy Sciences Network (ESnet), Lawrence Berkeley National Laboratory
Berkeley, CA 94720
Email: ctracy@es.net

***Abstract*—High-rate, large-sized (α) flows are of interest to providers for various reasons, e.g., they have the potential to degrade service quality for real-time flows, and users are sensitive to the throughput variance of these flows. In this paper, we present characteristics, such as size, duration, average rate, of α flows computed from NetFlow records collected over a 7-month period from 4 ESnet routers. Flows moving datasets as large as 811 GB and at rates as high as 5.7 Gbps were observed. Some source-destination pairs were found to repeatedly create α flows. An analysis of the rates of the 1596 repeated α flows created by one pair showed considerable variance, with minimum rate of 100 Mbps, maximum rate of 536 Mbps, and a coefficient of variation of 30%.**

I. Introduction

Problem statement and motivation: The objective of this work is to characterize high-rate large-sized flows from measurements obtained on a US-wide backbone network. These flows are referred to as α flows [1].

Core research-and-education networks (RENs), such as US Department of Energy (DOE)'s Energy Sciences Network (ESNet)[1] and Internet2[2], offer network services to DOE laboratories (labs) and universities, respectively, where scientific computing applications are executed. Large-sized datasets created by these applications are moved at high data rates between computing and storage locations [2].

There is interest in characterizing α flows for various applications, such as traffic engineering α flows [3], identifying BGP misconfigurations that could cause α flows to enter on less-desirable peering links, and identifying performance issues such as high throughput variance.

Solution approach: Our algorithm for size/duration computation is similar to that used by Fioreze et al. [4] with one difference. We aggregate records into flows for only those flows that had at least one NetFlow record in which a size threshold of 1 GB was exceeded (since size computation from sampled NetFlow records is not accurate for smaller-sized/lower-rate flows). This filtering step speeds up computation by reducing the size of the processed NetFlow records dataset. Our algorithm was applied to NetFlow records procured from ESnet for a characterization of high-rate large-sized (α) flows.

[1]http://www.es.net/
[2]http://www.internet2.edu

Findings/contributions: *First*, we characterized the size, duration and average rate of α flows observed at four ESnet routers over a 7-month period. *Next*, we compared the characteristics of α flows observed at different types of routers, and found certain patterns. For example, there were more α flows in the download direction from DOE labs than in the upload direction to DOE labs (which is consistent with the fact that most university scientists use the DOE supercomputing centers to run their applications and then download datasets from these centers). To study *persistency*, we determined the number of flows created by each source-destination IP address pair. The maximum number of α flows (flows that exceed 5GB in size and 100Mbps in rate) for a single source-destination pair was 1596, of which 75% experienced less than 167 Mbps while the highest rate was 536 Mbps. Such information is useful for initiating diagnostics to improve performance.

After reviewing related work in Section II, we present our size-duration-rate estimation algorithm in Section III. Section IV presents characteristics of high-rate large-sized flows observed at four routers of ESnet, and discusses potential network management applications. Our conclusions are presented in Section V.

II. Related work

Flow characterization studies include early work by Lan and Heidemann [5], above-described paper by Fioreze et al. [4], and an analysis of elephant flows [6]. Bayesian methods have been proposed for identifying high-rate and/or large-sized flows but not characterizing flow sizes and rates [7].

III. Algorithm for estimating flow size/rate

A. Terminology

Flow: A *flow* is defined to consist of all packets arriving with the same 5-tuple values {source IP address, destination IP address, source port number, destination port number, protocol type} with no consecutive inter-packet gaps greater than some fixed time threshold τ. The five tuples constitute the *flow IDentifier (flow ID).*

NetFlow records: A NetFlow record r is represented as

$$\{\omega_r, f_r, l_r, v_r, o_r\} \tag{1}$$

where ω_r is the (5-tuple) flow identifier, f_r is the Coordinated Universal Time (UTC) timestamp of the first packet in the record, l_r is the UTC timestamp of the last packet in the record, v_r is the number of packets in the record, and o_r is the cumulative number of octets (bytes) in the record.

Types of NetFlow records: A size threshold H is used to divide a day's set of NetFlow records collected from a router into two subsets: `Large` and `Small`. Since the duration of a NetFlow record r is upper-bounded by the active timeout interval a, the flow needs to have sent more than H bytes within the a-sec period following the first-packet arrival time (f_r) for the NetFlow record r to qualify as `Large`.

Types of flows: We define a β flow as a flow that has only `Small` NetFlow records, and a γ flow as a flow that has at least one `Large` NetFlow record. Thus, a γ flow may have multiple `Large` and `Small` NetFlow records. Since TCP varies its sending rate, not all NetFlow records of a γ flow will exceed the size threshold H. An α flow is defined to be a γ flow whose size and rate exceed specified (configurable) thresholds.

B. Algorithm

TABLE I: Notation

i	per-day index
j	flow-identifier (ID) index
k	γ-flow index
r	NetFlow-record index
$\mathbf{F}_i$	set of NetFlow records
$\mathbf{A}_i$	set of `Large` NetFlow records (size $> H$)
$\mathbf{W}_i$	set of unique flow IDs ω_r for records $r \in \mathbf{A}_i$
$\mathbf{B}_i$	set of `Small` NetFlow records r whose flow IDs $\omega_r \in \mathbf{W}_i$
$\mathbf{C}_{ij}$	set of NetFlow records r, s.t. $\omega_r = j$, for $j \in \mathbf{W}_i$
$\mathbf{E}_{ijk}$	Subset of $\mathbf{C}_{ij}$: records of a single γ flow
N_{ij}	number of γ flows
S_{ijk}	size of γ flow
D_{ijk}	duration of γ flow
ρ	packet sampling rate (e.g., 1/1000)
T_s	size threshold for α flows
T_r	average rate threshold for α flows

Using the notation in Table I, the steps for determining the size, duration and average rate of α flows are listed below:

1) From each day's set of NetFlow records, $\mathbf{F}_i$, determine sets $\mathbf{A}_i$, $\mathbf{W}_i$, and $\mathbf{B}_i$ using the size threshold H.
2) For each day i, the set $\mathbf{A}_i \bigcup \mathbf{B}_i$ is divided into disjoint subsets, $\mathbf{C}_{ij}$, $1 \le j \le |\mathbf{W}_i|$.
3) Order the record in each set $\mathbf{C}_{ij}$ by sorting on the first-packet timestamp (earliest-to-latest). The ordered set of records are $r_1, r_2, \cdots, r_{|\mathbf{C}_{ij}|}$.
4) Divide each set $\mathbf{C}_{ij}$ into disjoint subsets $\mathbf{E}_{ijk}$, $1 \le k \le N_{ij}$ such that a consecutive set of NetFlow records $\{r_n, r_{n+1}, \cdots, r_{n+u}\} \in \mathbf{E}_{ijk}$ iff

$$\begin{aligned} f_{r_{m+1}} - l_{r_m} &\le \tau \qquad n \le m < n+u \\ f_{r_n} - l_{r_{n-1}} &> \tau \qquad \text{for } n \ne 1 \\ f_{r_{n+u+1}} - l_{r_{n+u}} &> \tau \qquad \text{for } n+u \ne |\mathbf{C}_{ij}| \end{aligned} \tag{2}$$

   There is potential for a small gap between l_r and f_{r+1} for two consecutive records r and $(r+1)$ because of packet sampling. Therefore, as long as this gap is less than a time-threshold τ, the consecutive NetFlow records are considered to belong to the same flow.
5) A γ flow $k, 1 \le k \le N_{ij}$, appearing on day i with flow-ID $\omega_j \in \mathbf{W}_i$, and consisting of NetFlow records $\{r_n \cdots, r_{n+u}\} \in \mathbf{E}_{ijk}$, is characterized by

$$\begin{aligned} \text{Size } S_{ijk} &= (\frac{1}{\rho}) \sum_{m \in \mathbf{E}_{ijk}} o_m \\ \text{Duration } D_{ijk} &= l_{r_{n+u}} - f_{r_n} \\ \text{Av. rate } R_{ijk} &= \frac{S_{ijk}}{D_{ijk}} \end{aligned} \tag{3}$$

6) An α flow is a γ flow whose size $S_{ijk} > T_s$ and average rate $R_{ijk} > T_r$.

IV. Application to ESnet traffic

We procured NetFlow records from four ESnet routers for a 7-month time period, May-Nov. 2011, and applied the algorithm described in Section III-B to characterize γ and α flows. The NetFlow active-timeout interval configured in the ESnet routers was 60 sec and packet sampling rate was 1-in-1000 packets. Routers `router-1` and `router-2` are ESnet provider-edge (PE) routers each connected to a single customer (DOE lab) network. `Router-3` is a core ESnet router connected to multiple REN peers, while `router-4` is an ESnet router with commercial peering links. The obtained NetFlow datasets were for flows arriving at each of these routers in the input direction of only inter-domain interfaces. Since ESnet does not offer transit service, all packets are either sourced or destined to ESnet's customers (DOE labs). The NetFlow records collected at `router-1` and `router-2` correspond to *downloads* from DOE lab machines, while NetFlow records collected at `router-3` and `router-4` are for *uploads* to DOE lab machines.

After presenting the results generated by applying our algorithm to the NetFlow data in Section IV-A, the implications of these findings are discussed in Section IV-B.

A. Results

Three sets of results are presented:

1) aggregate characteristics of γ flows and α flows
2) statistics about three characteristics: size, rate, and duration, of γ flows and α flows, and
3) number of α flows as a function of the size and rate thresholds.

Aggregate characteristics of γ flows (H was set to 1 GB and τ was set to 60 sec) and α flows (using a size threshold T_s of 5 GB and rate threshold T_r of 100 Mbps) at each of the

routers across the observation period of 214 days are listed in Table II. The second row lists the number of unique source-destination pairs that generated γ flows, in the 214-day period. The third row represents the maximum number of per-day γ flows corresponding to a single γ flow ID. Multiple γ flows could have resulted from a TCP connection being held open for a long duration with gaps between flows. The last two rows present aggregate information about α flows.

TABLE II: Aggregate data on γ and α flows; across 214 days

	Routers, `router-`			
	1	2	3	4
No. of γ flows	28685	27963	2516	212
No. of unique /32 src-dst pairs gen. γ flows	1479	1611	193	158
Max. no. of per-day per-flow ID γ flows	33	56	6	1
No. of α flows	916	9538	986	16
No. of unique /32 src-dst pairs gen. α flows	95	419	89	14

Statistics for three characteristics of γ flows: size, rate, and duration, are presented in Tables III, IV, and V. These tables are independent, e.g., the largest-sized flow is not the same as the highest-rate flow. Skewness is defined as μ_3/σ^3, where μ_3 is the third moment and σ is the standard deviation.

TABLE III: Size in MB of γ flows; across 214 days

	Routers, `router-`			
	1	2	3	4
Min	1001	1001	1005	1010
1st Qu.	1149	1540	4050	1203
Median	1275	2869	4360	1532
Mean	2513	9046	17540	3612
3rd Qu	1701	8768	21380	3772
Max	633300	811600	233600	112800
skewness	25.35	12.56	2.37	10.09

TABLE IV: Average rate in Mbps of γ flows; across 214 days

	Routers, `router-`			
	1	2	3	4
Min	11.7	3.6	34.6	49.2
1st Qu.	160.9	147	117.6	130.9
Median	199.3	181.9	132.6	156.4
Mean	245.2	230.9	159	182.7
3rd Qu.	258.9	252.1	159.2	195.8
Max	5154	5757	979	776
skewness	7.36	3.95	3.82	2.86

TABLE V: Duration in sec of γ flows; across 214 days

	Routers, `router-`			
	1	2	3	4
Min	4.2	8.0	9.5	12
1st Qu.	41.8	60.9	190.9	54.9
Median	54.2	121.1	272	94.3
Mean	122.8	414.2	1098	235.6
3rd Qu.	73.6	398.9	1169	227.6
Max	32460	31910	13940	9978
skewness	23.76	10.33	2.32	10.99

Table VI presents results from a sensitivity analysis of the number of α flows to the size (T_s) and rate (T_r) thresholds.

TABLE VI: Sensitivity to size-rate threshold: No. of α flows

size	rate	Routers, `router-`			
		1	2	3	4
5GB	200Mbps	496	4475	201	3
10GB	200Mbps	357	2443	92	0
50GB	200Mbps	19	505	28	0
80GB	500Mbps	0	20	0	0

B. *Discussion*

The results presented in the previous section are discussed below in three groupings. *First*, we discuss the numerical values themselves to understand the range of sizes, rates, durations, and frequencies, of γ flows and α flows. *Next*, we compare the characteristics of flows observed at the different routers. *Finally*, an example application is described to demonstrate usage of this characterization of α flows.

Numerical values:

Most γ flow IDs have only single γ flows in a given day, but there are a few occasions when multiple γ flows have been observed on the same day for a given γ flow ID. As many as 56 γ flows were observed for a single five-tuple ID in one day (at `router-2`) as shown in Table II.

Across the 214-day period, of all the flows observed at the four routers, the largest-sized flow was 811.6 GB (max row of Table III) and the highest-rate flow enjoyed an aggregate rate of 5.76 Gbps (max row of Table IV), both of which were downloads passing through PE router `router-2`. The largest-sized flow had a rate of 301 Mbps, and the fastest flow size was 7.14 GB. The longest flow lasted 32460 sec (more than 9 hours) passing through `router-1`, during which time 370 GB was moved (max row of Table V).

At the lower end, rates as low as 3.6 Mbps were observed, also at `router-2`. This particular γ flow moved 1.9 GB, which means it lasted about 4181 sec (more than an hour).

The 99.9% value for size of γ flows that entered `router-1` from its connected DOE lab is 229.73 GB, which implies that only 28 flows were in the size range (229.73 GB, 633.3

GB). Similarly, the 99.9% rate value for γ flows passing through `router-2` was still less than 1 Gbps (even though the maximum rate for this router was 5.76 Gbps). This implies that only 27 flows out of the 27963 observed γ flows enjoyed (average) rates higher than 1 Gbps during the 7-month period.

Table VI shows that the number of α flows falls quickly as the size-rate threshold is increased, which is to be expected. Nevertheless, the absolute numbers are interesting to note. Router `router-2` connects ESnet to a supercomputing center, which explains that even at the high per-flow thresholds of 80 GB and 500 Mbps, 20 α flows were observed.

Comparison between flows observed at different routers:

As seen in Table II, there were many more γ flows in downloads from DOE labs than uploads to DOE labs (since downloads were observed at `router-1` and `router-2`, while uploads were observed at `router-3` and `router-4`). Also, more source-destination pairs engaged in transfers larger than 1 GB for downloads than uploads.

As seen in Tables III and IV, γ flows for downloads from DOE labs were larger in size and higher in rate. Uploads to DOE labs, observed at `router-3` and `router-4` were considerably slower, with the maximum per-flow average rate reaching only 776 Mbps at the commercial peering router `router-4` and only 979 Mbps at the REN-peering router `router-3`. Maximum flow sizes were also smaller. Table V shows that the longest downloads were longer than the longest uploads, but most γ flows are short in duration.

A comparison of the number of α flows across the 4 routers from Table VI shows a difference between the two PE routers. While `router-1` is a PE router connected to large national DOE lab, the significant research projects at this lab are in a single science discipline. In contrast, PE router `router-2` connects to a national scientific supercomputing center that is used by scientists from many disciplines. This explains the larger numbers of α flows for `router-2` when compared to `router-1` as seen in Table VI.

The maximum number of γ-flow and α-flow uploads per source-destination pair for `router-3` were 1229 and 325, respectively, while at `router-2`, the numbers for γ-flow and α-flow downloads per source-destination pair were 2913 and 1596, respectively. The maximum γ-flow and α-flow downloads per source-destination pair at `router-1` were 2860 and 445, respectively. Less than 10% of the source-destination pairs generated large numbers of repeated γ flows and α flows, which makes it somewhat easier for operators to provide better services (higher rates, lower variance) for these particular source-destination pairs.

Example application:

We considered the source-destination pair that generated the largest numbers of γ flows and also the largest number of α flows across the 214-day period. Since all these flows were between the same source and destination, and there were no network upgrades during the data-collection period, the bottleneck link rate and round-trip time were approximately the same, and all flow sizes are greater than 1 GB, which means TCP's Slow Start period could not have had a major influence on the average rate. Nevertheless, in the 2913 γ-flow set (observed at `router-2`), 75% of the flows experienced less than 161.2 Mbps while the highest rate experienced was 1.1 Gbps (size: 3.5 GB). Similarly, in the 1596 α-flow set (observed at `router-2`), 75% of the flows experienced less than 167 Mbps, while the highest rate experienced was 536 Mbps (size: 11 GB). Such information would allow the provider to initiate diagnostics to determine the causes of this throughput variance.

V. Conclusions

This paper presented size, duration, and average rate characteristics of high-rate, large-sized (α) flows, which were computed from NetFlow records collected over a 7-month period from 4 ESnet routers. Numerical values for the largest observed flow size, highest per-flow average rate, and longest duration are 811 GB, 5.7 Gbps, and 9 hours, respectively. A comparison of observed α flows across the four routers showed that downloads from DOE labs were more frequent, larger in size, and higher in rate, than uploads to DOE labs. Another characteristic that was quantified is a measure of persistency. The maximum number of α flows observed between the same source and destination was 1596, with a minimum per-flow average rate of 100 Mbps, and a maximum average rate of 536 Mbps. This example illustrates one of the potential network management applications for α-flow characterization, i.e., the identification of performance problems.

VI. Acknowledgment

The University of Virginia work was supported by the U.S. Department of Energy (DOE) grant DE-SC0007341 and NSF grants ACI-1340910, OCI-1127340, CNS-1116081, and CNS-1405171. The ESnet work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. DOE under Contract No. DE-AC02- 05CH11231.

References

[1] S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-level analysis and modeling of nework traffic," in *ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001, pp. 99–104.

[2] Z. Liu, M. Veeraraghavan, Z. Yan, C. Tracy, J. Tie, I. Foster, J. Dennis, J. Hick, Y. Li, and W. Yang, "On using virtual circuits for GridFTP transfers," in *The International Conference for High Performance Computing, Networking, Storage and Analysis 2012 (SC 2012)*, Nov. 10-16, 2012, pp. 81:1–81:11.

[3] T. Jin, C. Tracy, M. Veeraraghavan, and Z. Yan, "Traffic engineering of high-rate large-sized flows," in *Proc. of IEEE 14th High Performance Switching and Routing (HPSR) 2013*, July 8-11 2013.

[4] T. Fioreze, L. Granville, A. Pras, A. Sperotto, and R. Sadre, "Self-management of hybrid networks: Can we trust netflow data?" in *Integrated Network Management, 2009. IM '09. IFIP/IEEE International Symposium on*, 2009, pp. 577–584.

[5] K-C. Lan and J. Heidemann, "A measurement study of correlations of internet flow characteristics," *Computer Networks*, vol. 50, no. 1, pp. 46–62, 2006.

[6] P. Megyesi and S. Molnár, "Analysis of elephant users in broadband network traffic," in *Advances in Communication Networking Lecture Notes in Computer Science*. Springer, 2013, vol. 8115, pp. 37–45.

[7] Y. Zhang, B. Fang, and Y. Zhang, "Identifying high-rate flows based on bayesian single sampling," in *2010 2nd International Conference on Computer Engineering and Technology (ICCET)*, vol. 1, 2010, pp. V1–370–V1–374.